

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-242867
(43)Date of publication of application : 11.09.1998

(51)Int.Cl. H03M 7/30
G10L 9/14
G10L 9/18
H03M 7/42

(21)Application number : 09-040404
(22)Date of filing : 25.02.1997

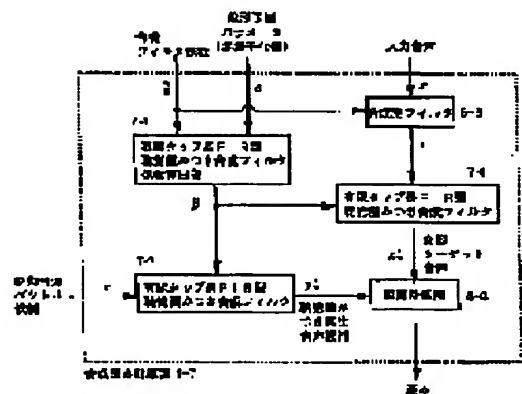
(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>
(72)Inventor : OOMURO NAKA
MANO KAZUNORI

(54) SOUND SIGNAL ENCODING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain a sound signal of high quality inexpensively at a low bit rate by adding an approximate error by an approximation filter for performing distortion calculation fast by discontinuing a tap of an FIR type composing filter halfway, etc., to an input sound, and using it as a target vector for code book searching.

SOLUTION: When finite-tap-length FIR filters 7-2 and 7-4 use taps as many as subframes, the encoding result matches that of a method which does not adopt approximation calculation, but the arithmetic process quantity becomes large. When the number of the taps is set to one, on the other hand, sufficient encoding quality can not be obtained because of distortion calculation between a driving sound source vector candidate and an ideal sound source vector. The number of taps is determined between 1 and subframe length (number of samples of subframe) in consideration of the balance between the encoding quality and arithmetic processing quantity. For 80 samples, there is no deterioration in auditory quality because of cancellation effect of composing filters 7-2 and 7-4 even when 2 to 6 taps are removed.



LEGAL STATUS

[Date of request for examination] 21.10.1998
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number] 3095133
[Date of registration] 04.08.2000
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-242867

(43) 公開日 平成10年(1998) 9月11日

(51) Int.Cl.⁹ 識別記号

H 0 3 M 7/30

G 1 0 L 9/14

9/18

H 0 3 M 7/42

F I

H 0 3 M 7/30

G 1 0 L 9/14

9/18

H 0 3 M 7/42

B

G

J

E

審査請求 未請求 請求項の数8 O L (全 15 頁)

(21) 出願番号 特願平9-40404

(22) 出願日 平成9年(1997) 2月25日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 大室 伸

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 間野 一則

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

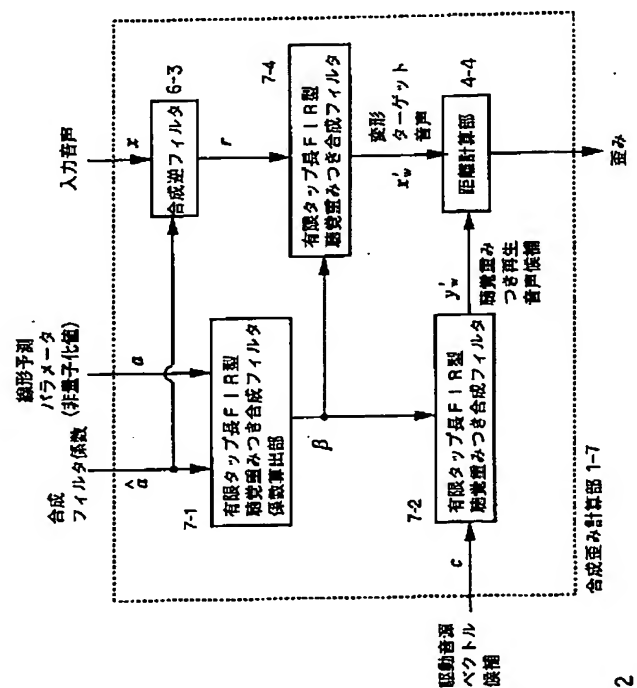
(74) 代理人 弁理士 草野 卓

(54) 【発明の名称】 音響信号符号化方法

(57) 【要約】

【課題】 符号帳から符号選択のための歪(距離計算)を少ないメモリ量で高速に行うことを可能とする。

【解決手段】 線形予測パラメータ a と、これを量子化した後、逆量子化した \hat{a} とを用いて、タップ長が2~6程度のFIR形聴覚重み付き合成フィルタ係数を算出し(7-1)、このフィルタ係数のFIRフィルタ7-2に駆動音源ベクトル候補を通す。係数 \hat{a} と対応する合成逆フィルタ6-3に入力音声を通し、その出力 r をフィルタ7-2と同一特性のフィルタ7-4に通して、フィルタ7-2で生じる近似誤差を付加し、これをターゲット音声として、フィルタ7-2の出力との距離計算を行い(4-4)、駆動音源ベクトルを探査する。



【特許請求の範囲】

【請求項1】 符号帳から取り出した時系列ベクトルより得られる駆動音源ベクトルにより、合成フィルタを駆動して音響信号を再生し、これと入力音響信号との間の歪が最小になるように上記符号帳を探索する符号化方法において、

上記合成フィルタを、高速に歪み計算ができるように簡略化した近似フィルタで表現し、

上記合成フィルタを上記近似フィルタで表現したことにもとづき生じる近似誤差を、上記入力音響信号に付加し、

この近似誤差が付加された入力音響信号を上記符号帳の探索に用いることを特徴とする音響信号符号化方法。

【請求項2】 請求項1に記載の音響信号符号化方法において、

入力音響信号を、合成フィルタの逆フィルタに通し、その逆フィルタ出力を、上記近似フィルタと同一の特性を持つフィルタに通すことによって上記近似誤差が付加された入力音響信号を得ることを特徴とする音響信号符号化方法。

【請求項3】 請求項1または請求項2に記載の音響信号符号化方法において、

上記近似フィルタを、有限タップ長のFIR型フィルタで実現することを特徴とする音響信号符号化方法。

【請求項4】 請求項3に記載の音響信号符号化方法において、

上記FIRフィルタのタップ長を、IIRフィルタを用いる場合と等価な結果が得られるタップ数に比べて少ない値に設定することを特徴とする音響信号符号化方法。

【請求項5】 請求項3または請求項4に記載の音響信号符号化方法において、

上記FIRフィルタのタップ長を、過去のサンプル点の値を1点用いる2タップ以上、過去のサンプル点の値を9点用いる10タップ以下の値に設定することを特徴とする音響信号符号化方法。

【請求項6】 請求項1から請求項5のいずれかに記載の音響信号符号化方法において、

上記符号帳として過去の駆動音源ベクトルよりなる適応符号帳と、固定符号ベクトルよりなる固定符号帳とを用い、

上記固定符号帳から出力される固定符号ベクトルを、ピッチに対応する周期で周期化してから駆動音源ベクトルとして用い、

固定符号ベクトルの探索の際には、上記固定符号ベクトルの周期化の周期の周期化逆フィルタを上記入力音響信号に通すと共に上記近似誤差を付加することを特徴とする音響信号符号化方法。

【請求項7】 請求項1から請求項6のいずれかに記載の音響信号符号化方法において、

上記近似フィルタを有限タップ長のFIR型フィルタで

表現し、その有限長FIRフィルタの係数から得られる、有限長のインパルス応答を用いてインパルス応答行列を作成し、

そのインパルス応答行列の転置行列と当該インパルス応答行列の積である、相関行列を計算してメモリに展開して蓄積し、

上記相関行列の値を参照しながら上記歪を計算することとを特徴とする音響信号符号化方法。

【請求項8】 請求項1から請求項7のいずれかに記載の音響信号符号化方法において、

上記近似フィルタは聴覚重み付きがなされていることを特徴とする音響信号符号化方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声、音楽などの音響信号の、スペクトル包絡特性を表すフィルタを音源ベクトルで駆動して音響信号を合成する予測符号化により、音響信号の信号系列を少ない情報量でデジタル符号化する高能率音声符号化方法に関する。

【0002】

【従来の技術】デジタル移動体通信において、電波を効率的に利用したり、音声または音楽蓄積サービス等で通信回線や記憶媒体を効率的に利用するために、高能率音声符号化方法が用いられる。現在、音声を高能率に符号化する方法として、原音声をフレーム（またはサブフレーム）と呼ばれる5～50ms程度の一定間隔の区間に分割し、その1フレームの音声を周波数スペクトルの包絡特性を表す線形フィルタの特性と、そのフィルタを駆動するための駆動音源信号との2つの情報に分離し、それぞれを符号化する手法が提案されている。この手法において、駆動音源信号を符号化する方法として、音声のピッチ周期（基本周波数）に対応すると考えられる周期成分と、それ以外の成分に分離して符号化する方法が知られている。この駆動音源情報の符号化法の例として、符号駆動線形予測符号化（Code-Excited Linear Prediction: CELP）がある。上記技術の詳細については、文献 M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates", IEEE Proc. ICASSP-85, pp.937-940, 1985に記載されている。

【0003】図8に上記符号化方法の構成例を示す。入力端子1-0に入力された音声 x は、線形予測分析部1-1において、入力音声の周波数スペクトル包絡特性を表す線形予測パラメータ a が計算される。得られた線形予測パラメータ a は線形予測パラメータ符号化部1-2において、符号化されて線形予測パラメータ復号部1-3に送られる。また、歪み計算に聴覚特性を考慮するなど、入力音声のスペクトル情報を利用して歪み計算を行う場合には、線形予測パラメータ a は歪み計算部1-6へも送られる。線形予測パラメータ復号部1-

3では、受け取った符号から合成フィルタ係数 a を再生し、合成フィルタ1-5に送る。歪み計算に聴覚特性を考慮する場合に、歪み計算部1-6において量子化前の線形予測パラメータ a を用いる代わりに、上記復号された線形予測パラメータ a を歪み計算に使用することもある。なお、線形予測分析の詳細および線形予測パラメータの符号化例については、例えば古井貞熙著“デジタル音声処理”(東海大学出版会)に記載されている。ここで、線形予測分析部1-1、線形予測パラメータ符号化部1-2、線形予測パラメータ復号部1-3および合成フィルタ1-5は非線形なものに置き換えてもよい。

【0004】駆動音源ベクトル生成部1-4では、1フレーム分の長さの駆動音源ベクトル候補 C を生成し、合成フィルタ1-5に送る。図9に駆動音源ベクトル生成部1-4の構成例を示す。適応符号帳2-1からは、バッファに記憶された直前の過去の駆動音源ベクトル(既に量子化された直前の1~数フレーム分の駆動音源ベクトル) $C(t-1)$ を、ある周期に相当する長さで切り出し、その切り出したベクトルをフレームの長さになるまで繰り返すことによって、音声の周期成分に対応する時系列ベクトルの候補 V_a が出力される。上記「ある周期」とは、歪み計算部1-6における歪み d が小さくなるような周期が選択されるが、選択された周期は、一般には音声のピッチ周期に相当することが多い。固定符号帳2-2からは、音声の非周期成分に対応する1フレーム分の長さの時系列符号ベクトルの候補 V_r が出力される。固定符号帳2-2には入力音声とは独立に符号化のためのビット数に応じてあらかじめ指定された数の候補ベクトルが記憶されている。適応符号帳2-1および固定符号帳2-2から出力された時系列ベクトルの候補は、乗算部2-4、2-5において、それぞれ重み符号帳2-3において作成された重み g_a 、 g_r が乗算され、これら乗算結果は加算部2-6において加算され、駆動音源ベクトルの候補 C となる。図2の構成例において、適応符号帳2-1を用いないで、固定符号帳2-2のみの構成としてもよく、子音部や背景雑音などのピッチ周期性の少ない信号を符号化するときには、

$$d = \| X_w - V_w \|^2$$

といった距離尺度を用いればよい。上記歪み尺度を最小にするような駆動音源ベクトルが選択される。図9に示したような駆動音源ベクトル生成の構成を用いる場合には、周期符号、固定符号、重み符号が決定される。なお、聴覚重みフィルタ4-2、4-3は、人間の聴覚特性を利用して再生音声の雑音感を低減するような歪み計算をするためのフィルタで、必ずしも用いる必要はない。

【0008】このとき、入力時系列音声ベクトル X は、入力音声信号そのままの場合もあるが、一般には、前サブフレームからの影響を差し引いた、時系列信号で

ビットを節約するために、適応符号帳2-1を用いない構成にすることも多い。

【0005】図8の説明に戻って、合成フィルタ1-5は、線形予測パラメータ復号部1-3の出力をフィルタの係数とする線形フィルタで、駆動音源ベクトル候補 C を入力として再生音声の候補 V を出力する。合成フィルタ1-5の次数すなわち線形予測分析の次数は、一般に10~16次程度が用いられることが多い。なお、既に述べたように、合成フィルタ1-5は非線形なフィルタでもよい。

【0006】歪み計算部1-6では、合成フィルタ1-5の出力である再生音声の候補 V と、入力音声 X との歪み d を計算する。この歪みの計算は、例えば聴覚重み付きなど、合成フィルタの係数 a または量子化していない線形予測係数 a を考慮にいれて行なうことが多い。図11に、聴覚重みづきを考慮して歪みを計算する構成例を示した。聴覚重みづきは、量子化していない線形予測パラメータ a もしくは量子化された合成フィルタ係数 a を用いた、聴覚重みフィルタの形で構成される。合成フィルタ4-1から出力される再生音声候補 V は、聴覚重みフィルタ4-2を通され、これは、同じく聴覚重みフィルタ4-3に通された入力音声との間で、歪み d が計算される。ここで、聴覚重みフィルタ4-2、4-3は通常同一のフィルタ係数を用いるため、聴覚重みフィルタ4-2、4-3は、距離計算部4-4の後に1つのフィルタとして入れても等価であるが、処理量の点から、図11に示されるように、距離計算部4-4の手前で2ヶ所に分けて入れることが多い。

【0007】この合成重み計算部1-7について更に述べると入力時系列音声ベクトル X は聴覚重みフィルタ4-3を通り、ターゲット音声 X_w となって、距離計算部4-4に送られる。一方、駆動音源ベクトル候補 C は、合成フィルタ4-1と聴覚重みフィルタ4-2を通り、聴覚重み付き再生音声候補ベクトル V_w となって、距離計算部4-4に送られる。距離計算部4-4では、ターゲット音声ベクトル X_w と再生音声候補ベクトル V_w の間の距離を測定する。このときの距離尺度には例えば、

$$(1)$$

あることが多い。また、図9に示したような駆動音源ベクトル生成の構成を用いる場合に、周期符号、固定符号、重み符号のすべての可能な組み合わせの中から、最適な組み合わせを1つ選択することは演算処理量の点から難しく、例えば周期符号、固定符号、重み符号の順に順次決定するか、途中で適宜候補を絞りながら順次探索し、最後に準最適な組み合わせに決定することが多い。このように順次決定または順次候補を残しながら探索する場合には、先に選択された符号ベクトル(例えば適応符号ベクトル)に起因する合成成分を入力音声から差し引き、駆動音源ベクトル候補 C には、これから決定し

たいベクトル成分のみ（例えば固定符号ベクトルのみ）を入力して歪み計算をする場合も多い。

【0009】図8において符号帳検索制御部1-8では各再生音声候補 \mathbf{v} と入力音声 \mathbf{x} との歪み d が最小となるような駆動音源符号を選択し、そのフレームにおける駆動音源ベクトルを決定する。なお、図9に示される適応符号帳2-1、固定符号帳2-2、重み符号帳2-3よりなる構成とする場合には、周期符号、固定符号および重み符号を選択し、これらを駆動音源符号とする。

【0010】符号帳検索制御部1-8において決定された駆動音源符号（周期符号、雑音符号、重み符号）と、線形予測パラメータ符号化部1-2の出力である線形予測パラメータ符号は、符号送出部1-9に送られ、利用の形態に応じて記憶装置に記憶されるか、または通信路を介して受信側へ送られる。図10に、上記符号化方法に対応する復号方法の構成例を示した。伝送路または記憶媒体から入力端子3-0に受信された符号のうち、線形予測パラメータ符号は線形予測パラメータ復号部3-2において合成フィルタ係数に復号され、合成フィルタ3-4および、必要に応じて後処理部3-5に送られる。駆動音源符号は、駆動音源ベクトル生成部3-3に送られ、符号に対応する音源ベクトルが生成される。なお、駆動音源ベクトル生成部3-3の構成は、図8に示された符号化方法の駆動音源ベクトル生成部1-4に対応する構成となる。合成フィルタ3-4は、駆動音源ベクトルを入力として、音声を再生する。後処理部3-5は、再生された音声の雑音感を聴覚的に低下させるような処理（ポストフィルタリングとも呼ばれる）を行うが、後処理部3-5は処理量削減等の関係から用いられないことも多い。

【0011】

【発明が解決しようとする課題】 CELP方式において問題となるのは、駆動音源ベクトル候補の選択をするための歪み計算に、非常に多くの演算処理が必要になることである。この問題に対して、Algebraic Code-Excited Linear Prediction (ACELP) という方式が提案されている。この方式は、固定符号帳を、フレーム長のベクトルパターンとして蓄えるのではなく、高さが1のパルスを、フレーム内に数本、例えば、40サンプルのフレームまたはサブフレームに対して、4本、適当な位置に立てることによって、固定符号ベクトルとする方式で、この駆動音源方式の採用と、歪み計算において演算順序を工夫することによって、従来の方式に比べて演算処理を大幅に減らすことができる。なお、ACELP方式の詳細は、例えば、文献、R. Salami, C. Laflamme, and J-P. Adoul, "8 kbit/s ACELP Coding of Speech with 10 ms Speech-Frame: a Candidate for CITTS standardization", IEEE Proc. ICASSP-94, pp. II-97に記載されている。また、同様の処理概念にもとづき、より高品質かつより低演算量の方法とし

て、この発明者等が既に出願した「音響信号符号化方法及び音響信号復号化方法」（特願平7-150550）がある。この方式では、固定符号ベクトルとして、高さが1のパルスのかわりに、隣接する2～数サンプルを単位とし、高さ情報を持つパルスパターンをフレーム内に配置する手法を用いることによって、より低演算量と高品質を両立している。

【0012】しかしながら、これらの方式においては、歪み計算に合成フィルタまたは聴覚重みづけフィルタ、またはそれらを合わせたフィルタを、インパルス応答またはFIR型のフィルタで表現することが多いが、フレームまたはサブフレームが長くなると、IIR型フィルタを用いる場合と等価な結果を得るためのFIRフィルタのタップ数が長くなり、演算量が従来方式に比べて逆に増加してしまうばかりでなく、歪み計算において計算の途中結果を格納するために、著しく大量のメモリが必要になるなどの問題がある。したがって、上記方法をそのまま、一般にサブフレームを長くする低ビットレート音声符号化に利用することは難しい。

【0013】一方図11の構成において、駆動音源ベクトル候補 \mathbf{c} を合成フィルタ4-1と聴覚重みづけフィルタ4-2に通す操作を、高速に実行するためには、これらの2つのフィルタを合わせて、等価なフィルタ特性を持つ1つの聴覚重み付き合成フィルタとするとよい。等価な1つのフィルタとするには、例えば合成フィルタ4-1の入力から聴覚重みフィルタ4-2の出力までのインパルス応答をフィルタ係数とすFIRフィルタで表現することができる。

【0014】図12は上記1つの等価なフィルタで表現する構成において、更に高速な歪み計算を実現する構成である。例えば、FIRフィルタ表現された聴覚重み付き合成フィルタを、有限タップで打ち切ったり、短いタップ数のARフィルタで近似したりして、あるいはFIRフィルタのタップ数を、IIRフィルタの場合と等価な結果を得るのに必要なタップ数よりも減らすなどの方法による厳密にはフィルタ特性の一致しない聴覚重み付き合成近似フィルタ5-2で代用する。これによって合成歪み計算における演算処理量およびメモリ量を減らすことができる。しかしながら、図12の構成を用いた場合、近似フィルタ5-2のフィルタ特性と、元の合成フィルタ4-1および聴覚重み付きフィルタ4-2の特性との差が大きくなると、近似誤差によって適当な駆動音源符号が選択されなくなり、再生音声の著しい品質劣化につながるため、事実上、サブフレームを長くすると、すなわちビットレートを低くすることは不可能であった。

【0015】この発明の目的は、低いビットレート、かつ安価なプロセッサで許容される範囲内の少ないメモリ量、少ない演算量で、高品質な再生音声を得られるような、音声または音楽などの音響信号をデジタル符号化

する方法を提供することにある。

【0016】

【課題を解決するための手段】この発明では、FIR型合成フィルタのタップを途中で打ち切るなどの高速に歪み計算ができるように簡略化した近似フィルタを合成歪み計算に用い、この近似フィルタで表現したことにもとづき生じる近似誤差を、入力音声に付加し、これを符号帳探索時のターゲットベクトルとする。

【0017】この構成により近似による影響を歪み計算において、相殺し、サブフレームの長い場合でも、非常に少ないメモリ量、処理量で、高品質な低ビットレート符号化方法を実現する。

【0018】

【発明の実施の形態】この発明の実施例を図1に示す。入力端子6-0よりの入力音声 x_a は、量子化された(復号された)合成フィルタ係数 a による合成フィルタの逆フィルタ(合成逆フィルタ)6-3を通り、理想の(量子化しない)駆動音源ベクトル r に変換される。 r は、図11において駆動音源ベクトル候補 c を入力とする合成フィルタ4-1に通したときに、入力音声 x との歪みがゼロになる理想の駆動音源ベクトルである。理想駆動音源ベクトル r は、聴覚重み付き合成近似フィルタ5-2と同じ特性の聴覚重み付き合成近似フィルタ6-4を通して変形ターゲット音声ベクトル x'_w となる。この時、聴覚重み付き合成フィルタ5-2で生じる近似誤差と同様の近似誤差が変形ターゲット音声ベクトル x'_w に付加されたものとなる。距離計算部4-4では、聴覚重み付き合成近似フィルタ5-2の出力である、近似誤差を含んだ聴覚重み付き再生音声候補 v'_w と、変形ターゲット音声ベクトル x'_w との間の距離を計算する。従ってこの距離計算においては聴覚重み付き合成近似フィルタ5-2で生じる近似誤差は、聴覚重み付き合成近似フィルタ6-4で付加された近似誤差と、距離計算の際に相殺され、歪み d (距離)を高い精度で計算できる。

【0019】図2は、図1におけるこの発明による方法において、合成近似フィルタ5-2、6-4を具体的に有限タップ長FIRフィルタ7-2、7-4の形で表現したものである。このときのタップ数は、サブフレーム長と同じ点数のタップ数を用いると、近似計算を用いない従来の方法と符号化結果が一致するが、演算処理量は多くなる。一方、タップ数を過去のサンプル値を用いない1タップ(これを0タップと呼ぶこともある)に設定すると、駆動音源ベクトル候補 c と理想駆動音源ベクトル r との間の歪みを、駆動音源レベルで測定する符号化方法になり、演算処理量は極めて少なくなるが、十分な符号化品質が得られない。タップ数は符号化品質と演算処理量のバランスを考慮して、1からサブフレーム長(サブフレームのサンプル数)の範囲で決定することになるが、この発明による方法では、サブフレームが例

えば80サンプルのときに、タップ数を2~6タップ程度まで減らしても、有限タップ長FIR型聴覚重み付き合成フィルタ7-2で生じる近似誤差が、ターゲット音声 x に対しても有限タップ長FIR型聴覚重み付き合成フィルタ7-4に付加されるため、実際の音声の符号化したときの信号対雑音比(SNR)、聴覚的品质とも、ほとんど劣化しないことを確認している。

【0020】図3は、駆動音源ベクトル生成部1-4の構成例において、固定符号ベクトル候補 v_r をピッチ周期化して用いる構成例である。前記ACELP方式や、「音響信号符号化方法及び音響信号復号化方法」

(特願平7-150550)でも図3に示す構成が用いられている。ピッチ周期化部8-7には、適応符号帳に入力される周期符号と同一の周期符号が入力され、周期符号に対応する周期で固定符号帳2-2の出力 v_r を周期化する。具体的な周期化操作は、固定符号ベクトル v_r に周期符号に対応するタップ位置のコムフィルタ(楕円フィルタ)をかけることが多い。またタップ位置は、整数サンプル位置でもよいし、非整数サンプル位置のコムフィルタを、アップサンプリングの手法を用いて実現してもよい。

【0021】図3の構成において、通常、適応符号帳8-1を探索するときには、固定符号帳2-2がないものとして最適な周期符号(または、歪みが小さくなる複数個の周期符号候補)を探索し、固定符号帳2-2を探索するときには、適応符号ベクトルを合成して得られる適応符号帳成分 v_a を、あらかじめ入力音声 x から除いたものを入力 x_r として、固定符号ベクトル v_r を合成して得られる成分 v_{rp} と x_r との間の歪みが最小になるような固定符号を探索するという手法が用いられる。この手法を用いる場合の、固定符号ベクトル合成歪み計算方法の構成例を図4に示す。図3におけるピッチ周期化部8-7は、乗算部2-5と順序を入れ替えることができるため、図4に示すように、乗算部2-5と合成フィルタ4-1の間にピッチ周期化部8-4を入れることができる。固定符号ベクトル v_r は乗算部2-5に送られる。乗算部2-5では v_r に重み g_r をかけて駆動音源ベクトル候補 c_r を生成し、ピッチ周期化部8-4に送る。 c_r はピッチ周期化された後、合成フィルタ4-1を通して再生音声候補 v_{rp} となり、聴覚重みフィルタ9-6を通して、距離計算部4-4に送られる。このとき、ピッチ周期化部8-4、合成フィルタ4-1、聴覚重みフィルタ4-2は3つのフィルタ特性を合成した特性を持つ1つのフィルタで表現すると、探索にかかる演算処理量を削減することできる。しかしながら、上記8-4、4-1、4-2の3つのフィルタの合成特性を持つフィルタをFIRフィルタで表現した場合、合成フィルタ9-5や聴覚重みフィルタ9-6の特性を持つFIRフィルタと違って、ピッチ周期に相当すると考えられる周期のタップ位置付近に大きな

値の係数を持つため、図2に示す構成例のように、短いタップ数でフィルタ係数を打ち切ってさらに高速な探索をすることができない。

【0022】この問題を解決し、ピッチ周期化のある場合でも高速に歪みを計算するこの発明の実施例を図5に示す。図5の構成例では、図1に示す構成例と同様に、図4における合成フィルタ4-1と聴覚重みフィルタ4-2の特性を合わせ持つフィルタを、聴覚重み付き合成近似フィルタ5-2に置き換える。図1の構成例と同様に、近似によって生じる歪みを入力側との間で相殺できるように、入力 x_r は合成逆フィルタ6-3を通し、フィルタ5-2と同じ特性の聴覚重み付き合成近似フィルタ6-4を通すが、この構成例では、図4におけるピッチ周期化フィルタ8-4の逆フィルタ（ピッチの周期性を取り除くフィルタ）10-4を、音声 x の入力側に入れる。この構成において、聴覚重み付き合成近似フィルタ5-2、6-4を、図2に示す構成例と同様に、有限タップ長FIR形聴覚重み付き合成フィルタで置き換えれば、非常に高速に符号帳の探索をすることができる。このときのFIRフィルタのタップ長は、図2の構成例と同様に、過去のサンプル値を用いない1タップ（0タップと呼ぶこともある）から、サブフレーム長までの間で、符号化品質と演算処理量とのバランスを考慮して決められるが、この発明による方法では、サブフレームが80点のときに、タップ数を2～6タップ程度ま

で減らしても、実際の音声を符号化したときの、信号対雑音比（SNR）、聴覚的品質とも、ほとんど劣化しないことを確認している。なお、図5の構成例において、合成逆フィルタ6-3、ピッチ周期化逆フィルタ10-4、聴覚重み付き合成近似フィルタ6-4が、すべて線形フィルタのときには、それらの順序を交換してもよい。

【0023】図6は、この発明による方法において、FIRフィルタを有限長で打ち切っても符号化音の品質劣化が非常に少ない利点を用いて、効率的に歪み計算を実施し、非常に高速な音声符号化を実現する構成例を示したものである。有限タップ長FIR型聴覚重み付き合成フィルタ係数算出部11-1では、合成フィルタ係数 a と量子化していない線形予測パラメータ a から、合成フィルタと聴覚重み付きフィルタを合わせた特性を持つ、聴覚重み付き合成フィルタをFIR型で実現したときのフィルタ係数を算出し、このフィルタ係数を有限タップ長で打ち切った係数 β を出力する。インパルス応答行列生成部11-2では、下記式（2）に示すように、FIRフィルタ係数を要素とする三角行列を生成する。ここで、Nはサブフレームのサンプル数を表す。式（2）において、係数 β は有限長で打ち切るため、例えば打ち切り次数をkとすると、 β_k から β_{N-1} までは0であって、式（3）のような行列となる。

【0024】

$$H = \begin{bmatrix} \beta_0 & 0 & 0 & 0 & \cdots & 0 \\ \beta_1 & \beta_0 & 0 & 0 & \cdots & 0 \\ \beta_2 & \beta_1 & \beta_0 & 0 & \cdots & 0 \\ \beta_3 & \beta_2 & \beta_1 & \beta_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{N-1} & \beta_{N-2} & \beta_{N-3} & \beta_{N-4} & \cdots & \beta_0 \end{bmatrix} \quad (2)$$

$$H = \begin{bmatrix} \beta_0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \beta_0 & 0 & 0 & \cdots & 0 \\ \beta_{k-1} & \vdots & \beta_0 & 0 & \cdots & 0 \\ & \beta_{k-1} & \cdot & \beta_0 & \cdots & 0 \\ & & \cdot & \vdots & \ddots & \vdots \\ 0 & & & \beta_{k-1} & \cdots & \beta_0 \end{bmatrix} \quad (3)$$

このとき上記行列の要素が0の部分は、メモリなどに記憶しておく必要がない。相関行列生成部11-3では、上記インパルス応答行列Hから、 $H^T H$ を計算する。このとき、係数の β_k から β_{N-1} までは0であるので、 $N \times N$ の行列計算をする必要がなく、 $k \times k$ の行列計算で $H^T H$ を求めることができる。例えば、kは、2から6の値に設定しても符号化音の品質がほとんど劣化しないため、 $N=80$ のときに、 80×80 の行列計算に比べて、例えば 5×5 の行列計算は著しい演算

処理量の削減となる。適応符号帳成分を除いた入力音声 x_r は、合成逆フィルタ6-3を通り、ピッチ周期化逆フィルタ10-4を通して、畳み込み部11-6に入力される。畳み込み部11-6では、ピッチ周期化逆フィルタ10-4の出力 r_p を、係数 β のFIRフィルタを通して、タップ打ち切り歪みを含むターゲット音声 x'_{rp} を求め、 x'_{rp} とH行列を、時間軸反転畳み込み操作もしくは行列演算によって、 $x'_{rp}^T H$ （ A^T は行列Aの転置を示す）を計算する。このと

きも打ち切り次数 k を小さくとていれば、非常に高速に計算を行うことができる。畳み込み部11-6は、別の手法を用いることもでき、相関行列計算部11-3の出力 $H^t H$ と、ピッチ逆周期化フィルタ11-5の出力 r_p から、行列演算によって、 $r_p^t (H^t H) r_p$ から、

を計算する。 d' は符号帳検索制御部に送られ、距離尺度 d' が最大になる(歪み尺度 d が最小になることと等価な)符号が選択される。

【0025】上述において、合成近似フィルタとしては必ずしも聴覚重み付き特性を与えたものでなくてもよい。なお特許請求の範囲において「フレーム」はフレームと、これを分割したサブフレームとの何れでもよい。

【0026】

【発明の効果】この発明の効果を確認するため下記の実験を行った。4.6 kbit/sのDual-Pulse CS-CELPを構成した。フレーム長は20ms、サブフレーム長は10ms(80点)とし、LPCの量子化はフレーム毎、その他はサブフレーム毎に行う。ビット配分はフレーム当り、LSP22ビット、適応符号8×2ビット、Dual-Pulse符号20×2利得符号7×2(計92(4.6 kbit/s))とし、Dual Pulseは、1サブフレームあたり3組配置し、位置11ビット、パタン6ビット、正負符号3ビットを割り当てる。

【0027】上記符号化器に実音声データを入力して、この発明方法の性能を調べた。音声データは、8 kHz サンプリングで、ITU-T G. 712帯域のフィルタをかけたものを用いた。図7に、FIRフィルタのタップを有限長で打ち切ったときの、打ち切りの次数とWSNRの関係を示した。WSNRは、最終的な合成音と入力音声との間で測定しているため、打ち切りのタップ数にかかわらず同一の尺度である。図中の方法(1)は、歪みを最小化するターゲット音声を従来の方法で求め、符号帳探索のためのフィルタのタップのみ打ち切った場合である。この場合は、20タップ以下になると急速に品質が劣化している。方法(2)はピッチ周期化逆フィルタを用いない図2に示したこの発明方法を適用した場合である。この方法を用いると、タップ数が2程度まではWSNRにほとんど変化がない。方法(3)はピッチ周期化逆フィルタを用いる図6に示したこの発明方法を適用した場合である。4.6 kbit/sのDual-Pulse CS-CELPは、Dual Pulseをピッチ周期化して駆動音源に用いるため、方法(3)を用いることによって非常に高速な符号化を実現できる。この場合の品質を方法(2)の場合と比較すると、全体的に0.3 dB程度低下しているものの、方法(2)の場合と同様に、タップ数を減らしてもWSNRはあまり低下しなかった。

H)を計算することもできる。このとき、上記 $X_{rp}^t H$ と $r_p^t (H^t H)$ は値が一致する。最終距離尺度計算部11-7では駆動音源ベクトル候補の固定符号帳成分 c_r と、 $H^t H \cdot X_{rp}^t H$ (または $r_p^t (H^t H)$)から、距離尺度

【0028】聴感上も6タップ程度使えば、全タップ使用する場合に比べてほとんど劣化が感じられない。また、方法(3)は方法(2)に比べてわずかに劣化が感じられる程度である。以上述べたようにこの発明によれば、非常に少ないタップ数で打ち切り、高速な符号帳探索、つまり高速な音声符号化を実現した場合でも、品質の劣化が非常に少ないことが確認された。

【図面の簡単な説明】

【図1】この発明における近似誤差を含んだ聴覚重み付き再生音声候補と、同じく近似誤差を含んだ変形ターゲット音声との間の距離を計算する方法の機能構成を示す図。

【図2】図1に示した方法において、聴覚重み付き合成近似フィルタを有限タップ長FIRフィルタの形で表現する例を示す機能構成図。

【図3】駆動音源ベクトル生成部の構成において、固定符号ベクトル候補をピッチ周期化して用いる機能構成例を示す図。

【図4】図3の構成を用いる場合の、固定符号ベクトル合成歪み計算方法の機能構成例を示す図。

【図5】図3に示すピッチ周期化のある場合に、この発明を適用し、ピッチ周期化逆フィルタを入力側に入れた歪み計算方法の機能構成を示す図。

【図6】この発明方法で、FIRフィルタを有限長で打ち切って効率的に歪み計算を実施し、非常に高速な音声符号化を実現する方法の機能構成例を示す図。

【図7】この発明を実際の音声符号化に適用した場合の、FIRフィルタタップの打ち切り次数とWSNRの関係を示すグラフ。

【図8】音声の符号駆動線形予測符号化(Code-Excited Linear Prediction: CELP)の機能構成例を示す図。

【図9】図8における駆動音源ベクトル生成部の機能構成例を示す図。

【図10】音声の符号駆動線形予測符号化(Code-Excited Linear Prediction: CELP)に対応する復号方法の機能構成例を示す図。

【図11】聴覚重みづきを考慮して歪みを計算する機能構成例を示す図。

【図12】従来の高速歪み計算方法の例で、聴覚重み付き合成フィルタの近似フィルタを合成歪み計算に用いる機能構成例を示す図。

【図1】

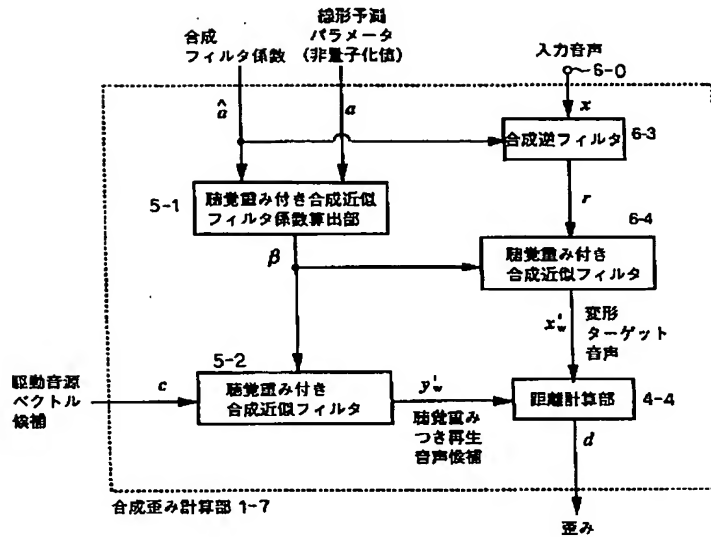


図1

【図7】

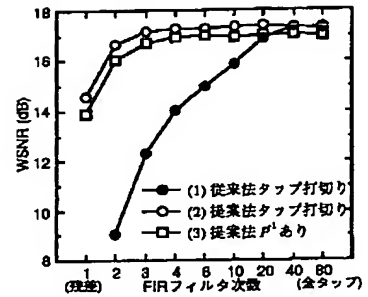


図7

【図2】

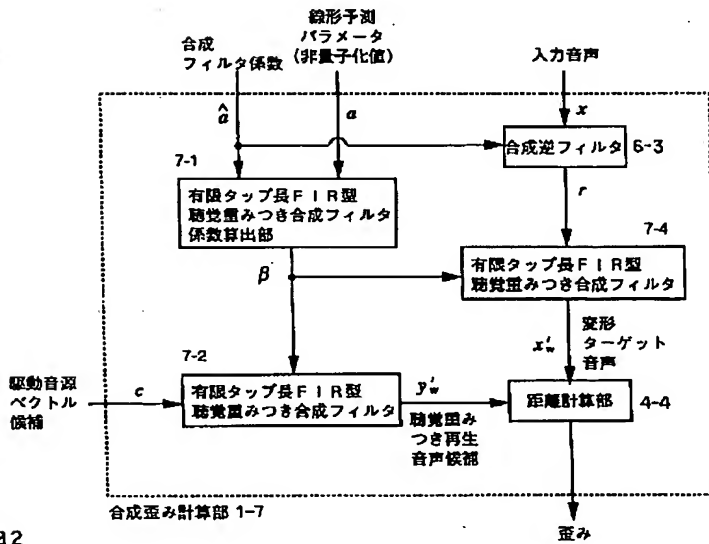


図2

【図3】

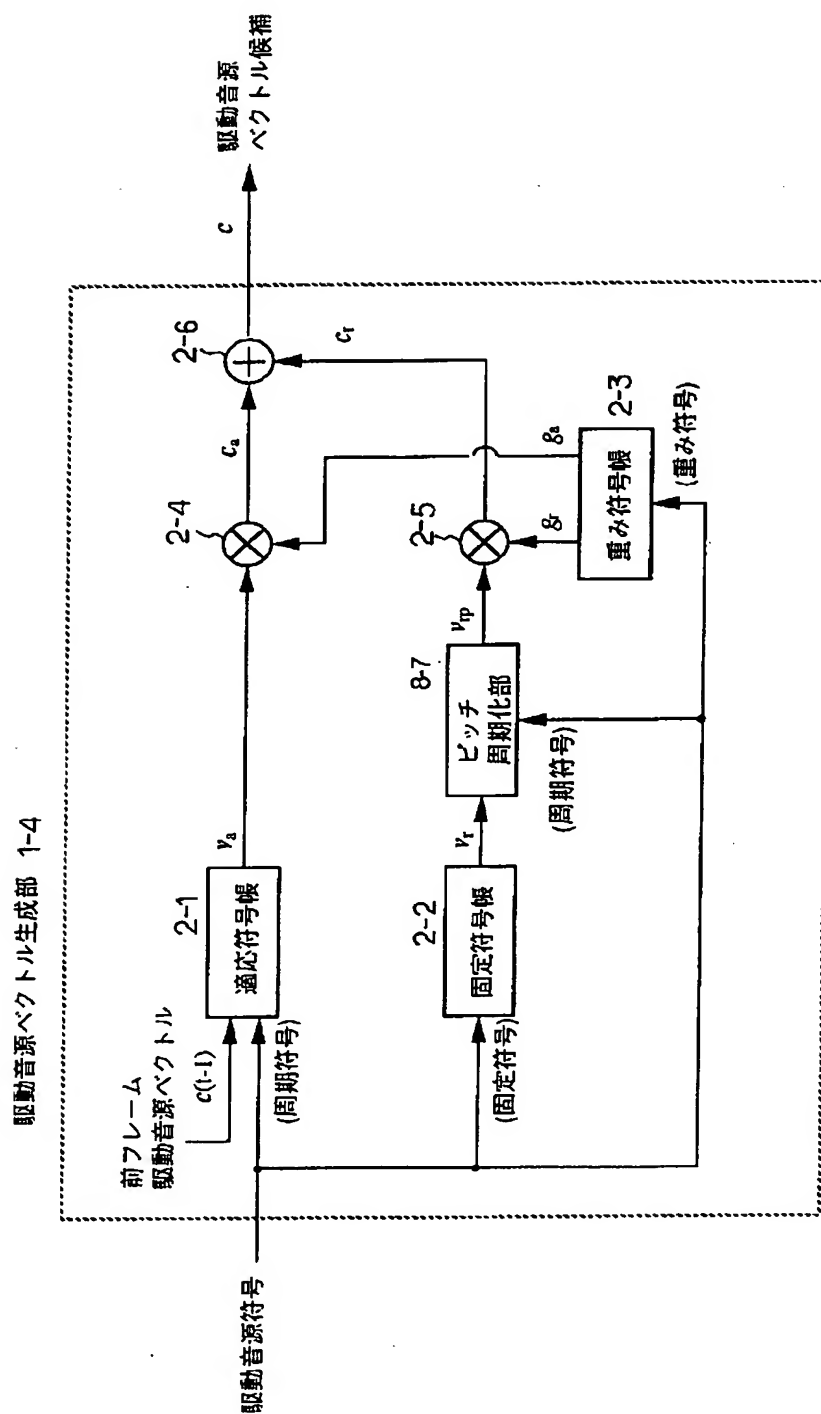


図 3

【図4】

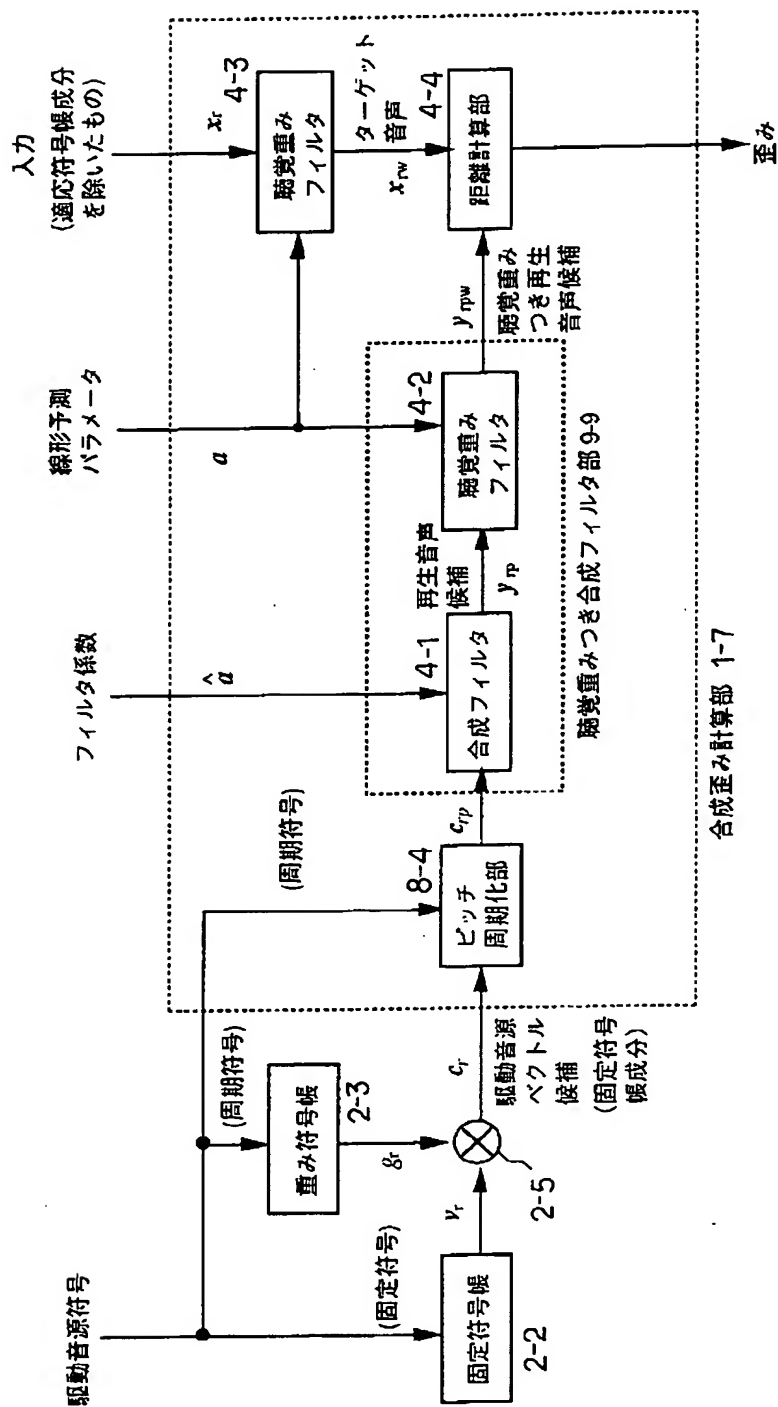


図4

【図5】

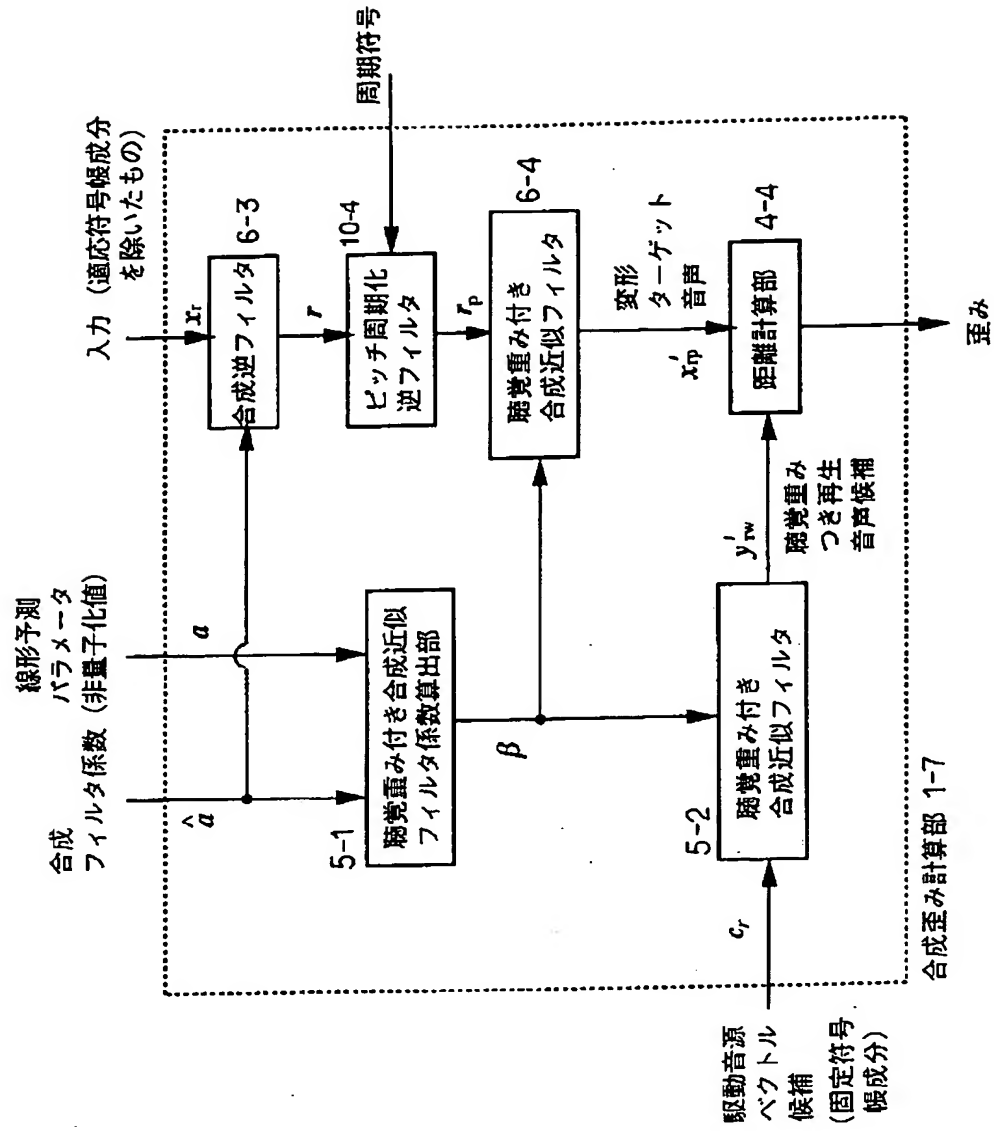


図5

【図6】

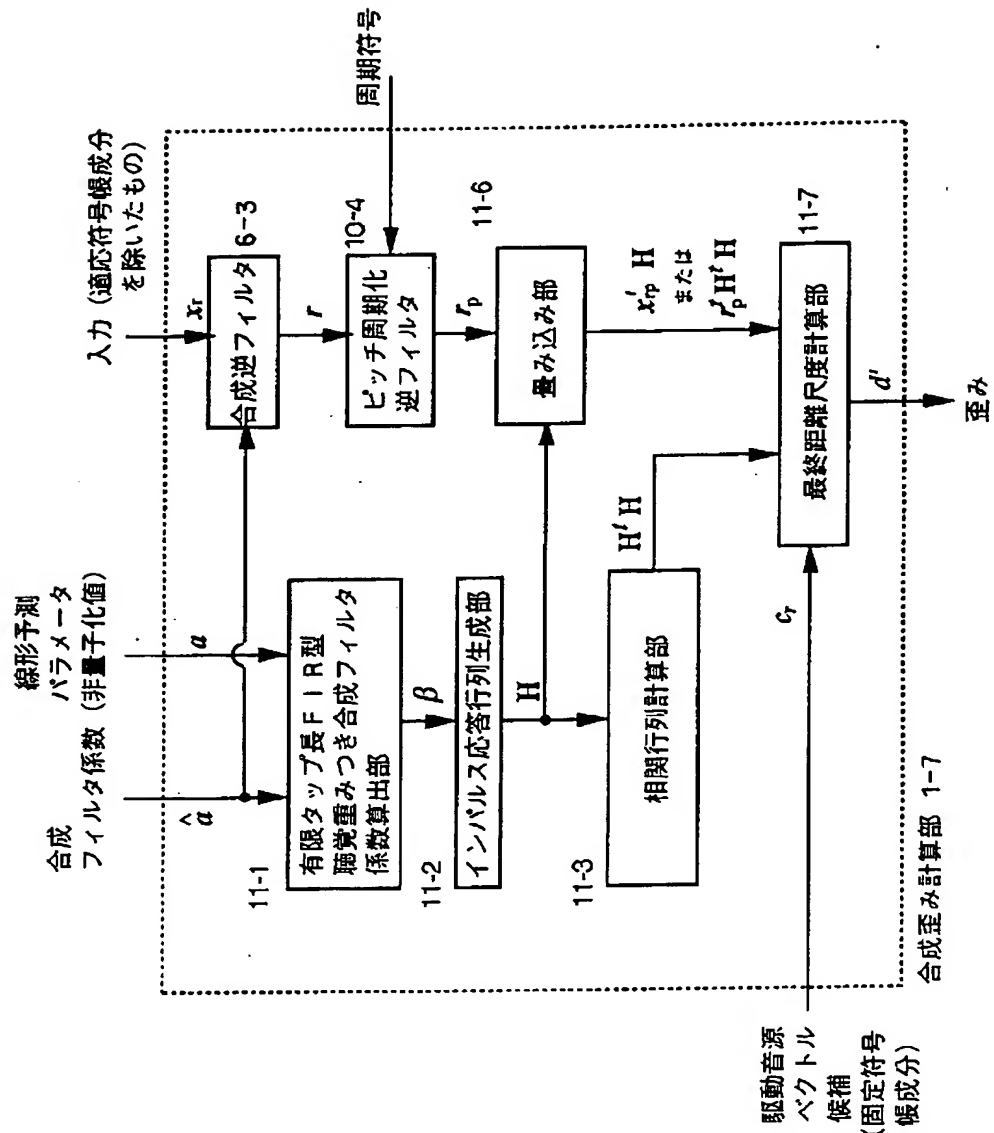


図 6

【図8】

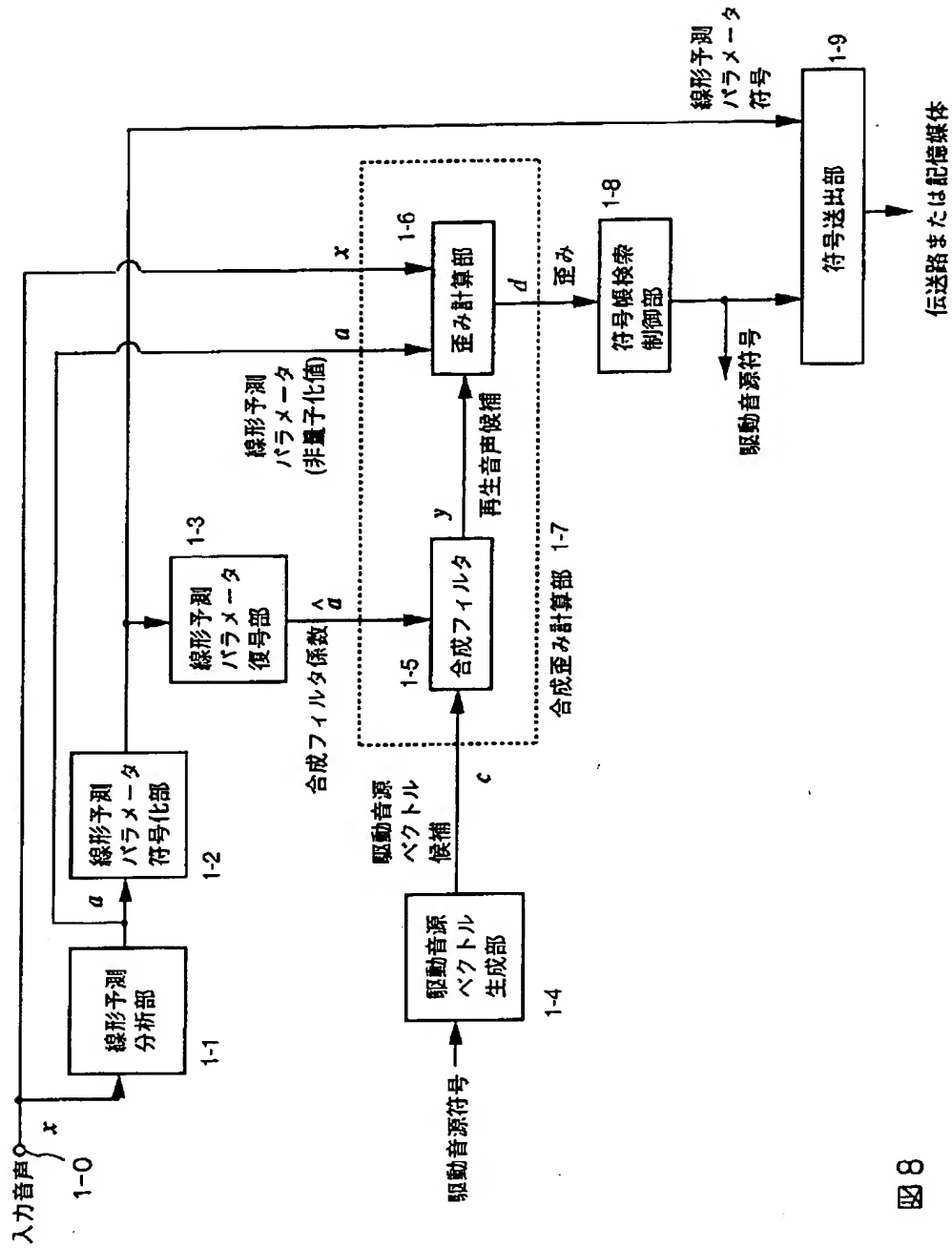


図 8

【図9】

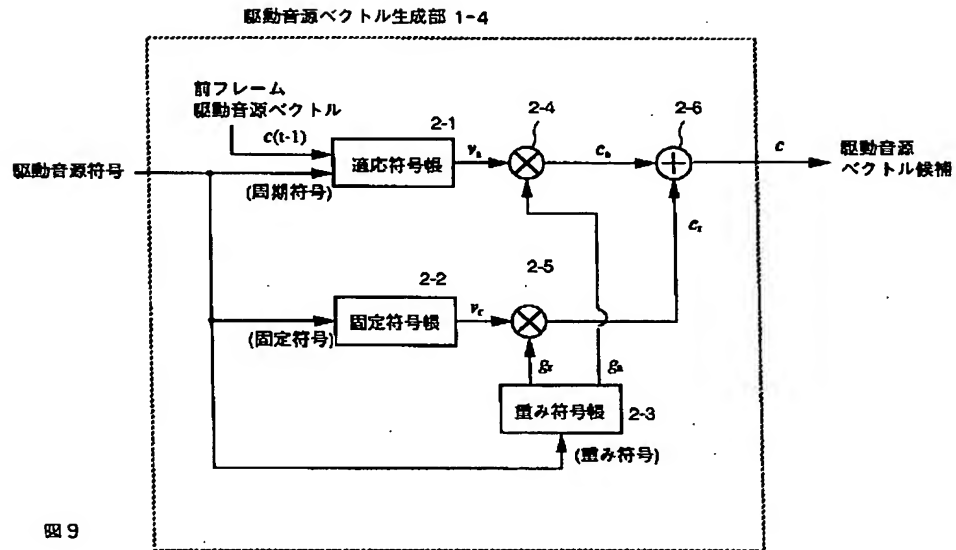


図 9

【図10】

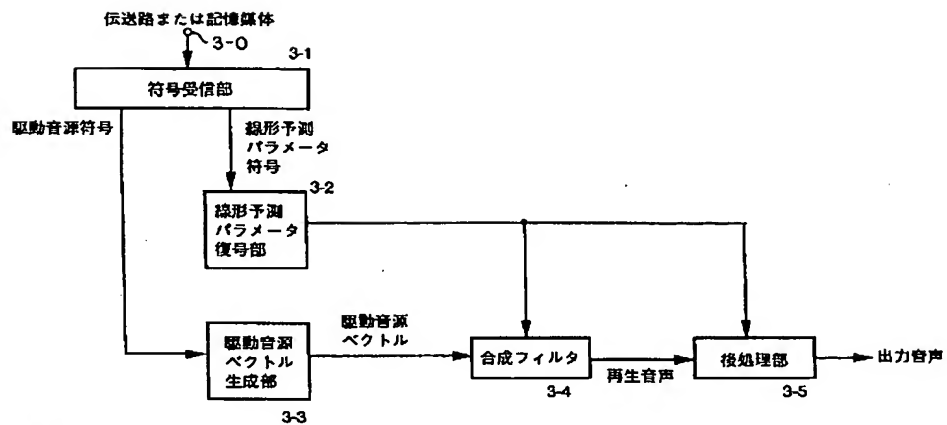


図 10

【図11】

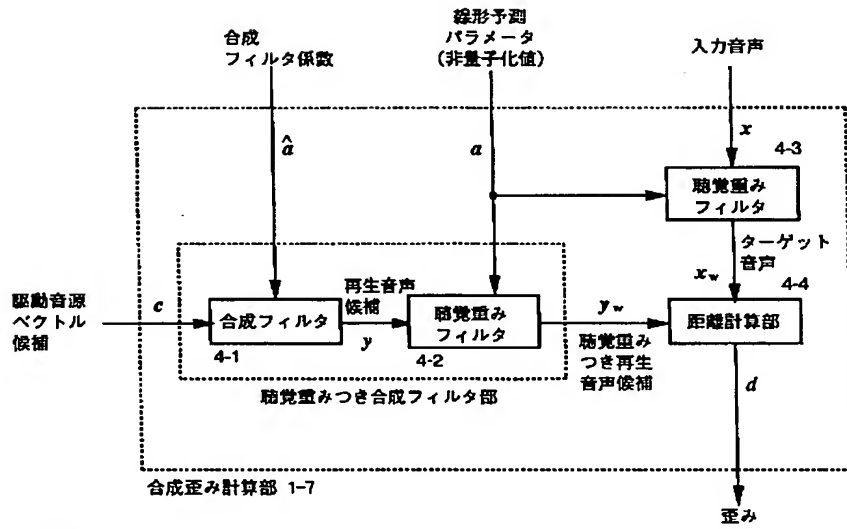


図 11

【図12】

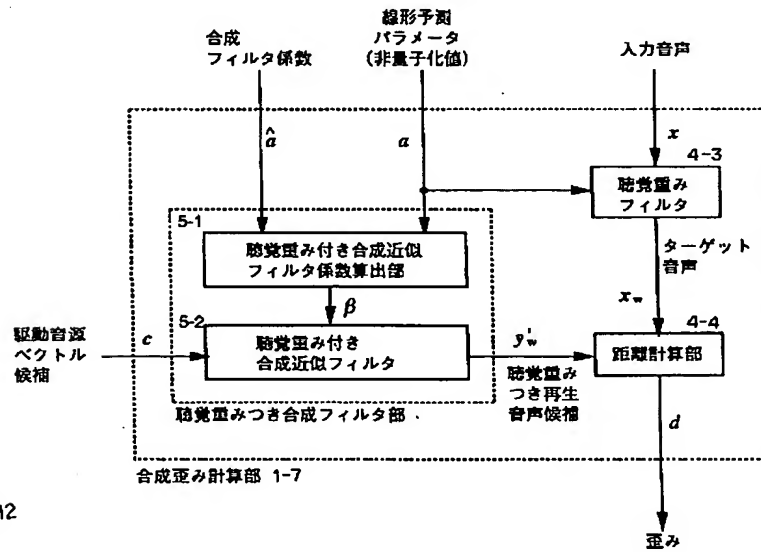


図 12